is the population size in the denominator, the standard error of the rate (not expressed per 1000 population) is $\sqrt{D}/P$. As D increases, the Poisson distribution rapidly approaches the normal distribution, and so the values for areas under a normal curve are frequently used to define confidence intervals around a rate. The 95 percent confidence interval around the rate D/P would therefore be $D/P \pm 1.96\sqrt{D}/P$. (The meaning of a confidence interval and the use of the normal distribution and 1.96 are explained in Part 1.)

Let us take the example of a death rate of 20 deaths (D) in a year out of a population of 5000 (P). The death rate is 20/5000 = .004 and the confidence interval is $.004 \pm 1.96 \sqrt{20}/5000$, or $.004 \pm .0018$. Expressed in terms of deaths per 1000 population, we can say that the "true" death rate for this population in this year was between 2.2 and 5.8, with 95 percent certainty. This is not very good precision, and unfortunately **many** rates for individual causes of death within single counties have rates with numerators less than 20. A useful rule of thumb is that any rate based on fewer than 20 events in the numerator may have a confidence interval that is wider than the rate itself. In the case above of a rate of 4.0 deaths per 1000 population with a numerator of 20, the width of the confidence interval is 3.6.

One way to reduce the standard error of a rate is simply to combine several years of data. Five-year rates, where deaths and population are added across five years in the numerator and denominator, are frequently shown in publications of the State Center for Health Statistics for this very reason. Another way to increase numerators is to combine geographic areas; for example, look at regional rather than county-level rates.

In many cases it is desirable to assess the statistical significance of a change in a rate over time, or of the difference between two rates in one period of time (for example between two geographic areas or population groups). The **standard error of the difference between two rates** is computed as

$$\sqrt{\frac{R_1^2}{D_1} + \frac{R_2^2}{D_2}}$$

where R_1 is the rate in area or period 1 with D_1 events (deaths for example) in the numerator, and R_2 is the rate in area or period 2 with D_2 events in the numerator. The difference between the two rates may be regarded as statistically significant at the 95 percent confidence level if it exceeds 1.96 standard errors of the difference as defined above. (*2*)

Suppose that the observed death rate for area 1 was 15.0 per 1000 population, with 20 deaths in the numerator, and the rate for area 2 was 20.0 per 1000 population with 10 deaths in the numerator. The difference between these rates is 5.0, but this is less than 1.96 standard errors of the difference which is

$$1.96\sqrt{\frac{(15.0)^2}{20} + \frac{(20.0)^2}{10}} = 14.0$$

With the number of deaths only 20 and 10, the difference between the two rates would have to be more than 14.0 in order to be significant at the 95 percent confidence level. The formula above will also work if the rates are expressed as proportions (not times 1000), i.e., .015 and .020, or as deaths per 100, per 100,000, etc.

One can use this formula for the standard error of the difference to solve for any unknown in the equation. For example, we may wish to know the level of statistical significance of the observed difference of 5.0. In this case we would solve for *Z* in the following equation, which is the number of standard errors that the difference of 5.0 represents:

$$Z\sqrt{\frac{(15.0)^2}{20} + \frac{(20.0)^2}{10}} = 5.0$$

$$Z\,(7.16) = 5.0$$

$$Z = .698\,.$$

From a table of areas under the normal curve we find that .698 standard errors corresponds to a probability of .484, which means that the above difference of 5.0 could occur due to chance about 48 times out of 100.

It is hoped that this discussion demonstrates that observed death rates should not be taken as exact measures of the true level of mortality in a population, and that, more generally, measures based on complete reporting from a population may have a substantial random error component.

REFERENCES

(*1*) Nathan Keyfitz, "Sampling Variance of Standardized Mortality Rates," *Human Biology*, vol. 38, 1966, p. 309-317.

(*2*) National Center for Health Statistics, *Vital Statistics of the United States, 1978*, Volume I - Natality, U.S. Department of Health and Human Services, Hyattsville, 1982; p. 4-15.

NC Division of Health Services
State Center for Health Statistics
7-84/500/5¢/copy